- 20 -

What is Claimed:

1. A method for generating structured document files from a document image, the method comprising the steps of:

segmenting the document image into one or more zones, at least one of the one or more zones containing a respective text image;

converting the respective text images within the at least one of the one or more zones to digital text;

automatically identifying layout information for each of the one or more zones;

labeling each of the one or more zones in accordance with a schema; and

automatically associating mark-up language tags with the labeled zones to generate the structured document files responsive to the identified layout information and a model file.

2. The method of claim 1, wherein the model file is associated with the schema and wherein the labeling step comprises at least the steps of:

automatically labeling each of the one or more zones responsive to the model file.

3. The method of claim 1, further comprising the steps of:

receiving editing commands corresponding to the one or more zones; and

updating the one or more zones responsive to the editing commands.

4. The method of claim 3, wherein the step of receiving editing commands includes the step of receiving text editing commands and the step of updating the one or more zones includes the step of editing the digital text responsive to the text editing commands.

5. The method of claim 3, wherein the step of receiving editing commands includes the step of receiving segmenting commands and the step of updating

3   the one or more zones includes the step of updating characteristics of the one or more
4   zones responsive to the segmenting commands.

1   6.   The method of claim 1, further comprising the step of:

2   receiving editing commands corresponding to the schema;

3   updating the schema responsive to the editing commands.

1   7.   The method of claim 1, wherein the respective text images are
2   displayed on a graphical user interface (GUI) and wherein the converting step comprises at
3   least the step of:

4   overlaying the respective text images displayed on the GUI with the at least
5   one of the one or more zones with the corresponding digital text.

1   8.   The method of claim 1, wherein the structured document files include
2   an XML file and an XSL file for each document image and wherein the generating steps
3   comprises at least the step of:

4   formating the XSL file such that information corresponding to each of the
5   labeled zones in the XML file is displayed in multiple layers on a web browser.

1   9.   The method of claim 1, wherein the steps of segmenting, converting,
2   labeling, and automatically associating mark-up language tags are performed sequentially
3   responsive to a selection of a workflow icon of a graphical user interface and wherein the
4   method further comprises the step of:

5   updating the workflow icon to represent a next step of the segmenting,
6   converting, labeling, and automatically associating mark-up language tags to be
7   performed, wherein the workflow icon presents a unique image corresponding to each
8   step.

1   10.   A system for generating structured document files from a document
2   image, the system comprising:

3   means for segmenting the document image into one or more zones, at least
4   one of the one or more zones containing a respective text image;

5   means for converting the respective text images within the at least one of
6   the one or more zones to digital text;

7   means for automatically identifying layout information for each of the one or
8   more zones;

9   means for labeling each of the one or more zones in accordance with a
10   schema; and

11   means for automatically associating mark-up language tags with the labeled
12   zones to generate the structured document files responsive to the identified layout
13   information and a model file.

1   11.   The system of claim 10, further comprising:

2   means for receiving editing commands corresponding to the one or more
3   zones; and

4   means for updating the one or more zones responsive to the editing
5   commands.

1   12.   The system of claim 10, further comprising:

2   means for receiving editing commands corresponding to the schema;

3   means for updating the schema responsive to the editing commands.

1   13.   A structured mark-up language generator for generating structured
2   document files from a document image, the generator comprising:

3   a document processor that: a) segments the document image into one or
4   more zones, at least one of the one or more zones containing a respective text image, b)
5   identifies layout information for each of the one or more zones, and c) converts the
6   respective text images within the at least one of the one or more zones to digital text;

7   a labeler that labels each of the one or more zones in accordance with a
8   schema; and

9   a structured document generator that generates the structured document
10   files responsive to the identified layout information and a model file.

1   14.   The generator of claim 13, further comprising:

2   an editor coupled to the document processor that enables editing of the
3   digital text and the one or more zones.

1   15.   The generator of claim 13, further comprising:

2   an editor coupled to the labeler that enables editing of the labels for each of
3   the one or more zones.

1   16.   A graphical user interface (GUI) for generating structured document
2   files from a document image, the GUI comprising:

3   a document panel for displaying a document image;

4   a schema panel for displaying a schema corresponding to the document
5   image; and

6   a workflow icon for directing the generation of at least one structured mark-
7   up language document from the document image, the workflow icon reflecting a next step
8   in a process to generate the at least one structured mark-up language document.

1   17.   The GUI of claim 16, wherein the process includes sequentially
2   performing the steps of loading an image, segmenting the image into zones, converting
3   text within the zones to digital text, labeling the zones, and generating the at least one
4   structured document and wherein the workflow icon is updated during the process to
5   present unique images corresponding to each step to be performed in the process.

1   18.   A computer readable medium including software that is configured to
2   control a general purpose computer to implement a method for generating structured
3   document files from a document image, the method comprising the steps of:

4   segmenting the document image into one or more zones, at least one of the
5   one or more zones containing a respective text image;

6   converting the respective text images within the at least one of the one or
7   more zones to digital text;

8   automatically identifying layout information for each of the one or more
9   zones;

10   labeling each of the one or more zones in accordance with a schema; and

11   automatically associating mark-up language tags with the labeled zones to
12   generate the structured document files responsive to the identified layout information and
13   a model file.

1   19.   The computer readable medium of claim 18, wherein the method
2   implemented by the software configured general purpose computer further comprises:

3   updating the one or more zones responsive to editing commands
4   corresponding to the one or more zones.

1   20.   The computer readable medium of claim 18, wherein the method
2   implemented by the software configured general purpose computer further comprises:

3   updating the schema responsive to editing commands corresponding to the
4   schema.